ARIC Manuscript Proposal # 2875

PC Reviewed: 10/11/2016 **Status: _____** **Priority:** 2
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title: **Gaseous Pollutants and DNA Methylation**

b. Abbreviated Title (Length 26 characters): **Gaseous Pollutants and DNAm**

2. Writing Group: WHI-EMPC & ARIC Epigenetics Working Groups
Writing group members: Interested members of the working groups with whom this proposal was discussed by conference call.

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. _KH__ **[please confirm with your initials electronically or in writing]**

First author: Dr. Katelyn Holliday
Address: UNC - Chapel Hill
Department of Epidemiology
Cardiovascular Disease & Environmental Programs
CVS Center, Suite 301-B
137 East Franklin Street
Chapel Hill, NC 27514
(T) 919-966-3168 or 1967
(F) 919-966-9800
Phone: 919-962-4962 Fax: 919-966-9800
E-mail: khausman@email.unc.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).
Name: Eric A Whitsel
Address: UNC - Chapel Hill
Departments of Epidemiology and Medicine
Cardiovascular Disease Program
CVS Center, Suite 301-B
137 East Franklin Street
Chapel Hill, NC 27514
(T) 919-966-3168 or 1967
(F) 919-966-9800

Phone: 919-9663168 Fax: 919-966-9800
E-mail: eric_whitsel@unc.edu

3. Timeline:Primary analyses & draft manuscript to be completed by late 2016

4. Rationale:

Gaseous air pollutants, including carbon monoxide (CO), oxides of nitrogen (NO_2/NOx), ozone (O_3), and sulfur dioxide (SO_2), have been linked to cardiovascular disease.[1] Primary pathways by which these pollutants influence health have been proposed, including inflammation/oxidative stress and autonomic nervous system imbalance.[1] Recently, DNA methylation has been proposed as a potential mechanism by which air pollution influences health.[2]

Methylation of DNA at *Cytosine-phosphate-Guanine* (CpG) sites serves as a normal process by which the body controls gene expression. Although methylation can persist over time and is thought to be heritable, it can also be modified through exposure to environmental pollutants. For example, methylation has been implicated as an effect measure modifier of air pollution-health associations[3-5] and exposure to air pollutants has been associated with atypical methylation in both animal[6] and human studies[7-11].

Currently, the limited literature examining associations between air pollution and DNA methylation has largely focused on particulate matter (PM) air pollution; however gaseous pollutants are also known to influence cardiovascular disease and may be acting through alterations of methylation patterns.[1] Current research is also geographically and sociodemographically limited, with much of it completed within a single study of elderly, largely Caucasian men from Boston, Massachusetts. Further, the air pollution exposures in these studies have typically been assigned using a single, central monitoring station. Therefore, discovery and replication of the relationship between gaseous pollutants and DNA methylation sites within a minority over-sample of Women's Health Initiative Clinical Trial (WHI CT) participants and the biracial Atherosclerosis Risk in Communities study (ARIC), will add significantly to this small body of literature. Further, the exposure assignments in the geographically diverse WHI CT and ARIC studies involved kriging and generalized additive mixed models, spatial and spatiotemporal methods that can estimate exposure at geocoded addresses of participants with reduced error when compared to simple assignment of centrally monitored exposures to distal locations.

5. Main Hypothesis/Study Questions:

To leverage the biracial, geographically diverse data within ARIC to replicate the associations between gaseous pollutants (CO, NO_2, NOx, O_3, and SO_2) averaged over 2, 7, 28 and 365 days and DNA methylation found within a minority-oversample of WHI CT participants.

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Overview. The general approach is to first conduct discovery analyses in WHI CT of gaseous pollutant-DNAm associations for each *Cytosine-phosphate-Guanine* (CpG) methylation site on the Illumina 450K Infinium Methylation BeadChip. The WHI CT analyses described herein will be based on DNA methylation data generated by WHI Ancillary Study #315 entitled, "*Epigenetic Mechanisms of PM-Mediated CVD Risk*" (WHI-EMPC; R01-ES020836; MPIs – Hou; Baccarelli; Whitsel). For each gaseous pollutant analysis, CpG sites will be ranked according to statistical significance followed by replication of CpG sites identified as significant or suggestive of significance within the Atherosclerosis Risk in Communities Study (ARIC), Cooperative Health Research in the Region Augsburg Study (KORA), and the Normative Aging Study (NAS). The replication analyses in ARIC will rely on air pollution data generated as part of the "*Modification of PM-Mediated Arrhythmogenesis in Populations*" ancillary study (MOPMAP; R01-ES017794; PI – Whitsel).

Study Population. The WHI CT AS #315 focuses on the core analytes subpopulation, an exam site- and race-stratified, randomly selected minority oversample of WHI CT participants who had repeated, fasting blood draws and resting, standard, twelve-lead electrocardiograms beginning at baseline. From this population, AS #315 randomly selected 2,200 participants with an available aliquot of DNA between 1993 and 2001 for DNA methylation assay, contemporaneous core analyte data, an address in the contiguous 48 U.S., and no conditions that affect the availability or accuracy of DNA methylation measures. The ARIC DNAm data are available from a subset of African American participants at visit 2/3 (N=2,905) and will soon be available for a subset of European American participants at visit 2/3 (N=1,102), allowing replication for the multi-ethnic WHI CT analyses.

Primary Outcomes. DNA methylation (DNAm) at CpG sites as determined by the Illumina 450K Infinium Methylation BeadChip, quantitatively represented by beta (the percentage of methylated cytosines over the sum of methylated and unmethylated cytosines), then quality controlled, batch-corrected, and normalized using Beta-MIixture Quantile (BMIQ) to correct for differences otherwise attributable to Type I and II probes.[12]

Main Exposure. Geocoded participant address-specific 2-, 7-, 28-, and 365-day mean concentrations of gaseous ambient air pollutants regulated under the Clean Air Act by the U.S. Environmental Protection Agency (EPA) according to its National Ambient Air Quality Standards (NAAQS) [carbon monoxide (CO), nitrogen oxides (NO_x), nitrogen dioxide (NO_2), ozone (O_3), sulfur dioxide (SO_2)]. Concentrations at the time of blood draw were estimated with a semiautomatic program using a spherical model to perform national-scale, lognormal ordinary kriging with the weighted least-squares method to estimate semivariograms in ArcView.[13] The model inputs were measurements from the US Environmental Protection Agency Air Quality System available from 1993 to align with date of participant blood draw, the monitor latitude and longitude, and geocoded addresses of WHI and ARIC participants. Gaseous pollutant measurements include a range of short and long term exposure averaging periods to facilitate future examination of gaseous pollutant-DNAm mediated effects on a variety of cardiovascular health endpoints with varying relevant exposure periods.

Covariates. Demographic covariates (age; center), technical covariates (plate; chip; row; column), Houseman estimates of cell type proportions (CD8-T, CD4-T, B cell, natural killer,

monocyte, and granulocyte), principal components for ancestral admixture, randomly assigned treatment group, relevant meteorological and seasonal covariates and potential confounders of interest (smoking status, alcohol use, body mass index, physical activity, individual education and neighborhood socioeconomic status).

Statistical Analyses.

Discovery Pollutant-DNAm Association Analyses in WHI CT. For each gaseous pollutant and exposure averaging period, covariate-adjusted, three-level, linear mixed effects longitudinal models will leverage repeated measures to estimate gaseous pollutant-DNAm associations. There will be a random intercept and slope for time at the participant level and for gaseous air pollutant at the WHI center level as well as a random intercept for technical covariates. These analyses will be stratified by race/ethnicity. Fixed-effects, inverse variance-weighted meta-analysis will be used to combine stratum-specific estimates in WHI.

Sensitivity Analyses. To assess sensitivity of WHI results, models will then be run on the entire WHI-EMPC population (i.e. without stratification by race/ethnicity) with control for race/ethnicity to include additional minority populations (i.e. American Indian or Alaskan Natives; Asian or Pacific Islanders; Others) that were too small to include in stratified analyses.

Replication Association Analyses in ARIC and other cohorts. For each gaseous pollutant and exposure averaging period, CpG sites identified as significant ($p<1.0x10^{-7}$) or suggestive ($p<1.0x10^{-5}$) for pollutant-related DNAm will be replicated within ARIC, KORA, and NAS, depending upon gaseous air pollutant data availability. Thresholds of significance for these analyses will be Bonferroni-corrected based on the number of CpG sites carried over for replication. Analyses will be of European Americans in NAS, Europeans in KORA, and stratified by race/ethnicity for ARIC (African American and European American). Fixed-effects, inverse variance weighting will again be used to meta-analytically combine the race/ethnicity-specific estimates.

Quality Control. We will follow established quality control protocols, including review of results by e.g. graphing the observed *P* values for each CpG site against CpG positions in Manhattan plots, by graphing them against the expected values from a theoretical χ^2 distribution in quantile-quantile (Q-Q) plots, and applying genomic control (λ) if inflation is observed.

Functional Annotation. Function annotation for the implicated CpG sites will be assessed using the UCSC Genome Browser[14] and the WashU Epigenome Browser[15] with data from the Encyclopedia of DNA elements (ENCODE)[16] and Roadmap Epigenomics Project.[17]

CONCLUSIONS

In this epigenome-wide association study, we will estimate associations between ambient gaseous air pollution and DNA methylation, the nature of which may ultimately affect our understanding of molecular consequences of exposure.

7.a. Will the data be used for non-CVD analysis in this manuscript? ____ Yes __x__ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = "CVD Research" for non-DNA analysis, and for DNA analysis RES_DNA = "CVD Research" would be used? ____ Yes ____ No
(This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? _x__ Yes ____ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = "No use/storage DNA"? __x_ Yes ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status.
ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/ARIC/search.php

__x___ Yes _____ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)? NA

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? _x__ Yes ____ No

11.b. If yes, is the proposal
_x__ A. primarily the result of an ancillary study (list number* _2009.08_)
___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____ _____)

*ancillary studies are listed by number at http://www.cscc.unc.edu/aric/forms/

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PubMed Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

13. Per Data Use Agreement Addendum, approved manuscripts using CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes __x__ No.

REFERENCES

1. Franklin BA, Brook R, and Pope CA. Air Pollution and Cardiovascular Disease. *Curr Prolb Cardiol.* 2015; 40:207-238.

2. Baccarelli A and Ghosh S. Environmental Exposures, Epigenetics and Cardiovascular Disease. *Curr Opin Clin Nutr Metab Care*. 2012; 15(4): 323-329.

3. Bind MA, Baccarelli A, Zanobetti A, et al. Air Pollution and Markers of Coagulation, Inflammation, and Endothelial Function: Associations and Epigene-Environment Interactions in an Elderly Cohort. *Epidemiology*. 2012; 23(2):332-340.

4. Lepeule J, Bind MA, Baccarelli A, et al. Epigenetic influences on associations between air pollutants and lung function in elderly men: the normative aging study. *Environ Health Perspect*. 2014;122:566-572.

5. Fu A, Leaderer BP, Gent JF, et al. An environmental epigenetic study of ADRB2 5'-UTR methylation and childhood asthma severity. *Clin Exp Allergy*. 2012; 42:1575-1581.

6. Ding R, Jin Y, Liu X, et al. Characteristics of DNA Methylation Changes Induced by Traffic-Related Air Pollution. *Mutation Research*. 2016; 796: 46-53.

7. Bind MA, Lepeule J, Zanobetti A, et al. Air Pollution and Gene-Specific Methylation in the Normative Aging Study. *Epigenetics*. 2014; 9(3): 448-458.

8. De Prins S, Kippen G, Jacobs G, et al. Influence of Ambient Air Pollution on Global DNA Methylation in Healthy Adults: A Seasonal Follow-Up. *Environment International*. 2013; 59:418-424.

9. Madrigano J, Baccarelli A, Mittleman MA, et al. Prolonged Exposure to Particulate Pollution, Genes Associated with Glutathione Pathways, and DNA Methylation in a Cohort of Older Men. *Environmental Health Perspectives.* 2011; 119(7):977-982.

10. Sofer T, Baccarelli A, Cantone L, et al. Exposure to Airborne Particulate Matter is Associated with Methylation Pattern in the Asthma Pathway. *Epigenomics*. 2013; 5(2): 147-154.

11. Salam MT, Byun HM, Lurmann F, et al. Genetic and epigenetic variations in inducible nitric oxide synthase promoter, particulate pollution, and exhaled nitric oxide levels in children. *J*

Allergy Clin Immunol. 2012;129: 232-239.

12. Wu MC, Joubert BR, Kuan PF, et al. A systematic assessment of normalization approaches for the Infinium 450K methylation platform. *Epigenetics*. 2014; 9(2):318-329.

13. Liao D, Peuquet DJ, Duan Y, et al. GIS Approaches for the Estimation of Residential-Level Ambient PM Concentrations. *Environmental Health Perspectives*. 2006; 114(9): 1374-1380.
14. Kuhn, Robert M., David Haussler, and W. James Kent. "The UCSC genome browser and associated tools." Briefings in bioinformatics (2012): bbs038.

15. Zhou, Xin, et al. "The human epigenome browser at Washington University." Nature methods 8.12 (2011): 989-990.

16. Rosenbloom, Kate R., et al. "ENCODE whole-genome data in the UCSC Genome Browser: update 2012." Nucleic acids research (2011): gkr1012.

17. Bernstein, Bradley E., et al. "The NIH roadmap epigenomics mapping consortium." Nature biotechnology 28.10 (2010): 1045-1048.